ON THE EM ALGORITHM FOR THE ESTIMATION OF SPEECH AR PARAMETERS IN NOISE

Marcin Kuropatwinski[1], *Bastiaan Kleijn*[2]

[1]VOICE LAB, Aleja Zwyciestwa 96/98, Gdynia, Poland
[2]Victoria University of Wellington, Victoria, New Zealand

ABSTRACT

In this paper, the estimation of speech AR parameters under noisy conditions is revisited. The EM algorithm serving this purpose was first proposed by Gannot et al. We present an extensive experimental study along with a new approach to implement the E-step of the algorithm. The new realization of the E-step uses matrix computations instead of a Kalman filter. By appropriate rearrangement of the E-step, the complexity $O(P(p+q)^3)$ of the Kalman filter approach has been reduced to $O(P\log P)$, where P is the frame length, p is the speech order and q is the noise order. In practice, a speed up of the E-step of at least two orders of magnitude has been achieved. An extensive evaluation of the algorithm shows that EM algorithm in its base form is unable to improve over a recent speech enhancement method proposed by Heusdens et al. and over an established Spectral Subtraction with Minimum Statistics method, as measured by various quality measures. However, with some modification it was possible to improve over these methods in terms of spectral distortion.

Index Terms— speech enhancement, estimation, EM algorithm, AR parameters

1. INTRODUCTION

Autoregressive (AR) modeling is ubiquitous in speech processing. It forms the basis for virtually all active speech coding standards. As the performance of coding degrades severely due to additive noise [1], algorithms that are able to estimate the speech AR coefficients from noisy signals have an important role to play. This motivates us to revisit the use of the expectation maximization (EM) algorithm for the estimation of the speech AR parameters from a noisy signal.

The estimation of the speech signal or its parameters from a noisy signal is a long- standing research problem. Its origin can be traced back to the seventies, cf. [2, 3]. During the eighties, the first attempts to use statistics trained from speech corpora during estimation were initiated, see, e.g., the paper by Porter [4]. From the beginning, there was an interest in estimating the autoregressive model (AR) parameters directly from noisy speech. An example is given by the paper by Lim [5]. Since then, the number of speech enhancement algorithms using the AR modeling has grown continuously. Algorithms have been proposed that carry out a maximum likelihood (ML) estimation via the EM algorithm [6], [7], search the codebooks to find the ML estimate [8] or perform the MMSE estimation of the AR parameters in the line-spectral frequencies representation [9], to mention only a few. The AR-parameter based approach to noise reduction remains an active research area. Some more recent efforts are focused on robustness under time-varying noise conditions, see e.g. [10]. Despite many years of development, all mentioned approaches suffer from a relatively high computational complexity, which makes the deployment on mobile platforms difficult.

The EM algorithm is an elegant, widespread method for statistical inference problems - for a complete discussion see [11, 12]. Though not studied in this paper, the EM algorithm for ML problems has a natural extension to maximum a-posteriori (MAP) inference. In that case it has an advantage over exhaustive search that in the M-step there is no need to search over a product space of speech and noise parameters. Another advantage, applying to both ML and MAP estimation, is that it does not require step-length setting procedure needed by gradient descent.

Our work is aimed at improving and benchmarking the EM algorithm approach to AR parameter estimation. Traditionally, the E-step is accomplished using a Kalman filter, cf. the classic paper by Gannot et al. [6]. Instead, we rearrange computations of the E-step into matrix computation and use a circulant matrix [13] based approximation to invert the matrices involved with the help of the FFT.

The remainder of the paper is organized into Signal Model Section, Algorithm Section, Experimental Section and Conclusions.

2. SIGNAL MODEL

In this section we describe the AR signal model, which provides a reasonable tradeoff between complexity and modeling accuracy. For this reason it is widely used in speech processing. Moreover, most broadband distortions can be accurately modeled as an AR process of low order and, thus, we use this model for the noise as well.

A vector process is observed:

$$\mathbf{r}=\mathbf{s}+\mathbf{n}\,, \tag{1}$$

where $\mathbf{s}$ and $\mathbf{n}$ are speech and noise independent random vectors describing a frame.

2.1 Likelihoods of speech, noise and noisy speech

The probability distribution of a speech frame (as well as a noise frame) is assumed to be multivariate Gaussian with zero mean and inverse of the covariance matrix given by $\mathbf{S} = \frac{1}{x^2}\mathbf{A}^T\mathbf{A}$ where $\mathbf{A}$ is a Toeplitz matrix with the first column $[\mathbf{a}^T, \mathbf{0}^T]^T$, $\mathbf{a} = [1, a_1, ..., a_p]^T$, where $\mathbf{0}^T$ is a zero vector of length $P - p - 1$, P is the frame length in samples, p is the model order and the $a_1, \ldots, a_p$ are the speech AR coefficients.

With the above assumptions, the likelihood of the noisy observation in a single frame can be shown to be:

$$p(\mathbf{r} \mid \theta) = (2\pi x^2 y^2)^{-P/2} \det(\mathbf{V})^{-P/2} \exp\left(\tfrac{1}{2}\mathbf{r}^T \left(\mathbf{N}^T\mathbf{V}^{-1}\mathbf{N} - \mathbf{N}\right)\mathbf{r}\right), \quad (2)$$

where $\mathbf{S} + \mathbf{N} = \mathbf{V} = \mathbf{Q}^T\mathbf{Q}$, and $\mathbf{N}$ is formed like $\mathbf{S}$ but contains the noise AR coefficients, y is the noise driving term standard deviation, $\theta = \{\theta_s, \theta_n\}$, where $\theta_s = [a_1, \ldots, a_p, x]^T$ and $\theta_n = [c_1, \ldots, c_p, y]^T$ are joint parameters of speech and noise respectively. The derivation of (2) is based on a multivariate convolution of speech and noise likelihoods.

3. ALGORITHM

In this section the proposed realization of the EM algorithm will be provided. In general terms the algorithm finds optimum of the function (2) defined in the previous section.

The objective of the proposed algorithm is to solve:

$$\hat{\theta} = \arg\max_{\theta} p(\mathbf{r} \mid \theta) \;. \quad (3)$$

We do not give theory of the EM algorithm here as it is easily found from a multitude of sources, e.g., [11, 12, 14], however, the computational details of our realization of this algorithm will be given in the sequel.

3.1 E-step

This step is analogous for speech and noise. The presentation below describes only the speech part of the E-step.

First we provide some Lemmas.

Lemma 1

$$\int_{R^P} \exp\left[\tfrac{1}{2}\left(\mathbf{x}^T\mathbf{x} + \mathbf{x}^T\mathbf{d}\right)\right] d\mathbf{x} = \left(2\pi\right)^{P/2} \exp\left(\frac{\mathbf{d}^T\mathbf{d}}{8}\right).$$

Lemma 2

$$\int_{R^P} \exp\left[\tfrac{1}{2}\left(\mathbf{x}^T\mathbf{x} + \mathbf{x}^T\mathbf{d}\right)\right] \mathbf{x}^T\mathbf{M}\mathbf{x}\, d\mathbf{x} = \left(2\pi\right)^{P/2} \exp\left(\frac{\mathbf{d}^T\mathbf{d}}{8}\right)\left(\mathrm{tr}\left(\mathbf{M}\right) + \frac{\mathbf{d}^T\mathbf{M}\mathbf{d}}{4}\right)$$

where $\mathbf{M}$ is an arbitrary square matrix. The proofs of the Lemmas are a standard exercise in multidimensional integration.

We recall that the so-called Q function, e.g., [11, 12, 14], is defined as:

$$Q(\theta_s \mid \hat{\theta}_{(k)}) = \mathrm{E}[\log(p(\mathbf{s} \mid \theta_s) \mid \mathbf{r}, \hat{\theta}_{(k)}], \quad (4)$$

where k is the EM iteration number.

The derivation of the Q function starts with:

$$\log[p(\mathbf{s} \mid \theta)] = C - P\log(x) - \tfrac{1}{2}\mathbf{s}^T\mathbf{S}\mathbf{s}, \quad (5)$$

where C is some constant. In the sequel matrices depending only on the conditioning argument $\hat{\theta}_{(k)}$ of the Q function will bear an apostrophe. It can be shown that, after omitting irrelevant terms, the following holds:

$$Q(\theta_s \mid \hat{\theta}_{(k)}) = -(A_1 + \tfrac{1}{2}A_2), \quad (6)$$

$$A_1 = \int_{\Omega_s} P\log(x)\exp\left\{-\tfrac{1}{2}\left[\mathbf{s}^T\mathbf{V}'\mathbf{s} - 2\mathbf{s}\mathbf{N}'\mathbf{r}\right]\right\} d\mathbf{s}, \quad (7)$$

$$A_2 = \int_{\Omega_s} \mathbf{s}^T\mathbf{S}\mathbf{s}\exp\left\{-\tfrac{1}{2}\left[\mathbf{s}^T\mathbf{V}'\mathbf{s} - 2\mathbf{s}\mathbf{N}'\mathbf{r}\right]\right\} d\mathbf{s}. \quad (8)$$

After substituting $\mathbf{s} = \mathbf{Q}'^{-1}\mathbf{h}$ (to diagonalize $\mathbf{V}' = \mathbf{Q}'^T\mathbf{Q}'$) we can apply *Lemma 1* to A_1 and *Lemma 2* to A_2:

$$A_2 = (2\pi)^{P/2}\exp\left(\frac{\mathbf{d}^T\mathbf{d}}{8}\right)\left(\mathrm{tr}\left(\mathbf{H}\right) + \frac{\mathbf{d}^T\mathbf{H}\mathbf{d}}{4}\right) = (2\pi)^{P/2}\exp\left(\frac{\mathbf{d}^T\mathbf{d}}{8}\right)\left(A_3 + A_4\right), \quad (9)$$

where: $\mathbf{M}^{-T} = (\mathbf{M}^{-1})^T$, $\mathbf{d} = -2\mathbf{Q}'^{-T}\mathbf{N}'\mathbf{r}$, $\mathbf{H} = \mathbf{Q}'^{-T}\mathbf{S}\mathbf{Q}'^{-1}$.

Applying the circulant approximation to Toeplitz matrices, cf. [9], we can continue derivation for A_3 and A_4. Since the trace of a matrix is invariant under cyclic permutations we get:

$$A_3 = \tfrac{1}{x^2}\mathrm{tr}(\mathbf{V}'\mathbf{A}^T\mathbf{A}). \quad (10)$$

Further, it can be shown that:

$$A_3 = x^{-2}\mathbf{a}^T\left(\mathbf{W}_0\tilde{v}_0 + \ldots + \mathbf{W}_{P-1}\tilde{v}_{P-1}\right)\mathbf{a}, \quad (11)$$

where

$\bar{\mathbf{a}} = [\mathbf{a}^T, \mathbf{0}^T]^T = [\bar{a}_0, ..., \bar{a}_{P-1}]^T$, $\tilde{\mathbf{a}} = \frac{1}{x^2}\left|\mathrm{DFT}\left(\bar{\mathbf{a}}\right)\right|^2 = [\tilde{a}_0, ..., \tilde{a}_{P-1}]^T$,

$\overline{\mathbf{c}} = [\mathbf{c}^T, \mathbf{0}^T]^T = [\overline{c}_0, \ldots, \overline{c}_{P-1}]^T$, where $\mathbf{c} = [1, c_1, \ldots, c_q]^T$ is vector of noise AR coefficients, $\tilde{\mathbf{c}} = \frac{1}{y^2}\left|\mathrm{DFT}\left(\overline{\mathbf{c}}\right)\right| = [\tilde{c}_0, \ldots, \tilde{c}_{P-1}]^T$.
Given our results thus-far it can be shown that:

$$\tilde{v}(\tilde{a}_i + \tilde{c}_i)^{-1} \quad . \tag{12}$$

For A_4 it holds:

$$A_4 = x^{-2}\mathbf{a}^T\left(\mathbf{W}_0\tilde{z}_0 + \ldots + \mathbf{W}_{P-1}\tilde{z}_{P-1}\right)\mathbf{a} \tag{13}$$

where, given: $\tilde{\mathbf{r}} = \frac{1}{y^2}\left|\mathrm{DFT}\left(\mathbf{r}\right)\right| = [\tilde{r}_0, \ldots, \tilde{r}_{P-1}]^T$, we have:

$$\tilde{z}_i = (\tilde{c}_i\tilde{v}_i)^2\tilde{r}_i \,, \tag{14}$$

and:

$$\mathbf{W}_i = \mathbf{w}_i^H\mathbf{w}_i \,, \tag{15}$$

$$\mathbf{w}_i = [\exp(-\, j\tfrac{2p}{P}i0), \ldots, \exp(-\, j\tfrac{2p}{P}ip)] \,, \tag{16}$$

where H denotes Hermitian transpose and $j = \sqrt{-1}$.

Therefore, finally, the speech part of the Q function reads:

$$Q(\theta_s, \hat{\theta}_{(k)}) = -\log(x) - \tfrac{1}{2}\mathbf{a}^T\mathbf{F}'\mathbf{a} \,, \tag{17}$$

$$\mathbf{F}' = \tfrac{1}{P}\sum_i \mathbf{W}_i(\tilde{v}_i + \tilde{z}_i) \tag{18}$$

3.2 M-step

In this subsection the M-step of our EM algorithm will be presented based on results of the E-step from the previous subsection. It can be shown that:

$$\hat{\mathbf{a}} = -\overline{\mathbf{F}}^{-1}\overline{\mathbf{f}} \,, \tag{19}$$

$$\mathbf{F}' = \begin{bmatrix} f & \overline{\mathbf{f}}^T \\ \overline{\mathbf{f}} & \overline{\mathbf{F}} \end{bmatrix} , \tag{20}$$

and that the thus-obtained forms the speech AR parameters for the $(k+1)$ iteration of the EM:

$$\mathbf{a} = [1, \hat{\mathbf{a}}^T]^T \,. \tag{21}$$

For the variance holds:

$$x^2 = \mathbf{a}^T\mathbf{F}'\mathbf{a} \,. \tag{22}$$

The noise AR parameters can be computed in an analogous manner.

3.3 Discussion

Our fast algorithm can be implemented by applying equation (18) for the E-step, and equations (19), (22), for the M-step, subsequently for each iteration. Compared to the conventional implementation that uses a Kalman filter to compute the $\mathbf{F}'$ matrix, FFT based computation of equation and its constituents results in a reduction of computational effort of Kalman filter approach. This is due to using the FFT as a main computational procedure. In our practical application, this means a savings of two orders of magnitude. The fast algorithm is useful for practical enhancement applications. It also facilitates the fast implementation of experiments. In section 4 we exploit this ability and provide insight in the performance of the EM estimation of the speech AR model coefficients from noisy signals.

4. EXPERIMENTS

In this section an extensive experimental study is presented. The speech and noise material used in this section has been taken from the NOIZEUS database. The NOIZEUS database is included in [15].

4.1 Database description

The database contains clean and noisy speech files. We used the narrowband part of the database. The experiments have been performed for the following conditions: multitalker babble, car interior, busy street, train. The algorithms have been tested for SNR 5, 10, 15 dB.

For each pair condition/SNR the database contains 30 utterances that contain all phones of English. The utterances are spoken by three male and three female subjects, each contributing five sentences. A detailed description of the recording equipment and other conditions during production of the database is contained in [15].

4.2 Algorithm settings

The speech AR model order was set ten throughout experiments. The frame length has been set to 200 samples at 8kHz. The frame advance was 100 samples and overlap-add with a Hanning window was used to form the enhanced signal. For the noise we tested orders q equal to two and four. For each file we performed 100 iterations of the algorithm. A numerical safeguard was used, which stopped iterations if unstable AR polynomial had been encountered as a result of the M-step. Initialization of the EM algorithm iteration was obtained from the Heusdens et al. speech enhancement system, cf. [16, 17].

4.3 Quality measures

For each enhanced utterance we computed a variety of quality measures, namely: Itakura-Saito distortion measure [18] (IS), likelihood ratio [18] (LLR), root mean spectral distortion [19] (SD), SNR [20], segmental SNR [20] (segSNR), PESQ measure [21] (PESQ), composite measure [22] (COM).

In [22] it was shown that the IS measure correlates best with the enhanced signal quality. For the spectral distortion we used two variants: a normalized measure (denoted SDn in the sequel), which compares only spectral shapes – irrespective of the driving term variances, which are set to one, and an unnormalized measure, which compares the unnormalized spectra (denoted SDu in the sequel).

Table 1 Spectral distortion measurements for Heusdens et al. and optimized EM algorithms

		SD normalized [dB] median		SD unnormalized [dB] median	
	SNR_{IN} [dB]	A2	EM optimized	A2	EM optimized
BABBLE	5	5.0	4.8	6.8	6.4
	10	3.7	3.4	5.0	4.4
	15	2.6	2.3	3.4	2.8
CAR	5	4.3	4.2	6.7	6.4
	10	3.6	3.3	5.0	4.4
	15	2.7	2.3	3.4	2.8
STREET	5	4.9	4.7	7.0	6.8
	10	4.0	3.8	5.5	5.0
	15	3.2	2.9	4.4	3.7

Table 2 Measurements for input SNR 5dB, first column contains environment, second algorithm and remaining columns various quality measures

		IS (median)	LLR (median)	SNR [dB]	segSNR (median) [dB]	PESQ [MOS]	COM [MOS] (sig;bak;ovl)
STREET	A1	1.12	0.64	8.5	0.2	2.2	3.1;2.3;2.6
	A2	1.54	0.79	9.0	1.2	2.3	2.8;2.2;2.4
	EM	1.45	1.00	8.7	1.2	2.0	2.4;2.1;2.1
CAR	A1	1.10	0.58	9.8	1.1	2.3	3.2;2.4;2.7
	A2	1.52	0.69	10.2	2.0	2.4	3.0;2.4;2.6
	EM	1.50	0.90	9.5	1.8	2.1	2.6;2.3;2.2

4.4 Experiments

To start, as a sanity check, we examined the likelihoods and confirmed that the likelihood increases consistently from iteration to iteration. We ran also some comparisons to the conventional, non-optimized EM which shows that the differences in estimated parameters are insignificant (on the order of 10^{-4} MSE).

Examining the data from experiments we note that the EM algorithm hardly improves over the first iteration. In the case of the SD, in rare cases, the performance increased also for the second and third iteration. The only quality measure that improves until the last iteration is the IS measure. This can be explained by the IS measure being closely related to likelihood, which is what is maximized by the EM algorithm. Some detailed results are contained in **Table 2**.

We compared the results of the EM algorithm to some competing algorithms. We have chosen two algorithms for comparison. The first one is the established and popular spectral subtraction algorithm as described the by Berouti [23], combined with the noise PSD estimator of Martin [24] (denoted A1 in the sequel) and the second is a recent, algorithm by Heusdens et al., [16], and Erkelens et al., [17], (denoted A2 in the sequel).

Summarizing qualitatively the results, both A1 and A2 yield generally better results in almost all measures than the plain EM algorithm. A bit surprising is that A2 consistently gets lower COM scores than A1, see **Table 2**. For presentation we have chosen the car and street noise – the tendencies for other noises are the same. We found that $q = 2$ performs better than $q = 4$.

When we optimize the number of iterations, then our EM algorithm performs better than A1 and A2 for spectral distortion. The optimal number of iterations of the EM depends (nonlinearly) on the frame SNR. The higher the SNR, the more iterations are beneficial. For low frame SNR it is optimal not to iterate EM at all, but for high frame SNR the optimal number of iterations reaches 30-40. Introducing a frame SNR dependent number of iteration decreases the spectral distortion by up to 0.4 dB of SDn and up to 0.7 dB of SDu over A2, with a larger change for the conditions with high SNR, see **Table 1**.

5. CONCLUSIONS

In this paper we elaborated on the classic enhancement algorithm first proposed by Gannot et al. [6]. We improved their method in terms of computational complexity. A significant acceleration of the E-step of two orders of magnitude was achieved by rearranging the E-step appropriately, with negligible loss in accuracy. The proposed improvement makes the real-time implementation of the EM-algorithm in low-cost devices a reality.

Our results are in general consistent with those by Gannot et al. However, the optimal number of iterations suggested in their paper was four. We found that setting the number of iterations globally constant is suboptimal. Better results were achieved by making the number of iterations dependent on the frame SNR. With this improvement our approach performs better than two competing algorithms in terms of Spectral Distortion.

Although the improvement over the state-of-the-art in terms of the chosen quality measure is not large, the proposed new realization of the E-step can bring advantages in future systems that require a-priori knowledge about speech AR parameters. Since there is still a large gap between SD values achieved by the plain EM algorithm and the results obtained earlier with usage of a-priori knowledge [9], we can expect significant gains in accuracy in future EM algorithms by utilizing a proper prior during estimation.

ACKNOWLEDGEMENTS

Marcin Kuropatwinski wishes to express his sincere gratitude to Mr. Jacek Kawalec for funding this work and to Mr. Tomasz Szwelnik for providing superb work environment.

6. REFERENCES

[1] M. Kuropatwinski, D. Leckschat, K. Kroschel, A. Czyzewski, and C. Hales, "Speech Enhancement for Linear-Predictive-Analysis-by-Synthesis Coders," in *Eurospeech'99 Proceedings*, 1999.

[2] J. Lim and A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proc. IEEE,* vol. 67, pp. 1586-1604, 1979.

[3] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 27, pp. 113-120, 1979.

[4] J. Porter, "Optimal Estimators for Spectral Restoration of Noisy Speech," in *ICASSP'84 Proceedings*, 1984, pp. 53-56.

[5] J. Lim and A. Oppenheim, "All-Pole Modeling of Degraded Speech," *IEEE Transactions Acoustics, Speech and Signal Processing,* vol. 26, pp. 197-210, 1978.

[6] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and Sequential Kalman Filter-Based Speech Enhancement Algorithms," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 6, pp. 373-385, 1998.

[7] B.-G. Lee, K. Y. Lee, and S. Ann, "An EM-based Approach For Parameter Enhancement With an Application to Speech Signals," *Signal processing,* vol. 46, pp. 1-14, 1995.

[8] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Speech Enhancement Using A-Priori Information," in *Interspeech'03 Proceedings*, 2003, pp. 1405-1408.

[9] M. Kuropatwinski and W. B. Kleijn, "Estimation of the Short Term Predictor Parameters under Noisy Conditions," *IEEE Transactions on Audio, Speech and Language Processing,* vol. 14, pp. 1645-1655, 2006.

[10] T. Rosenkranz and H. Puder, "Improving Robustness of Codebook-Based Noise Estimation Approaches with Delta Codebooks," *IEEE Transactions on Audio, Speech and Language Processing,* vol. 20, pp. 1177-1188, 2011.

[11] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*: Wiley, 2007.

[12] T. K. Moon, "The Expectation-Maximization Algorithm," *IEEE Signal Processing Magazine,* vol. 13, pp. 47-60, 1996.

[13] P. J. Davis, *Circulant Matrices*. New York: Wiley, 1979.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society,* vol. 39, pp. 1-38, 1977.

[15] P. C. Loizou, *Speech Enhancement: Theory and Practice*: CRC press, 2013.

[16] R. C. Hendriks, R. Heusdens, and J. Jensen, "MMSE Based Noise PSD Tracking with Low Complexity," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 4266-4269.

[17] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum Mean-Square Error Estimation of Discrete Fourier Coefficients with Generalized Gamma Priors," *IEEE Transactions Audio, Speech, and Language Processing,* vol. 15, pp. 1741-1752, 2007.

[18] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality*: Prentice Hall Englewood Cliffs, NJ, 1988.

[19] K. K. Paliwal and W. B. Kleijn, "Quantization of LPC Parameters," *Speech Coding and Synthesis,* pp. 433-466, 1995.

[20] J. H. Hansen and B. L. Pellom, "An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms," in *ICSLP*, 1998, pp. 2819-2822.

[21] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual Evaluation of Speech Quality (PESQ)-a New Method for Speech Quality Assessment of Telephone Networks and Codecs," in *ICASSP'01 Proceedings*, 2001, pp. 749-752.

[22] Y. Hu and P. C. Loizou, "Evaluation of Objective Measures for Speech Enhancement," in *Interspeech Proceedings*, 2006.

[23] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," in *ICASSP'79 Proceedings*, 1979, pp. 208-211.

[24] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Speech and Audio Processing,* vol. 9, pp. 504-512, 2001.